

ENDOGENOUS TREATMENT EFFECTS FOR COUNT DATA MODELS WITH ENDOGENOUS PARTICIPATION OR SAMPLE SELECTION

MASSIMILIANO BRATTI^a and ALFONSO MIRANDA^{b,*}

^aDepartment of Economics, Business and Statistics, Università degli Studi di Milano, Milan, Italy

^bDepartment of Quantitative Social Science, Institute of Education, University of London, London, UK

SUMMARY

In this paper, we propose an estimator for models in which an endogenous dichotomous treatment affects a count outcome in the presence of either sample selection or endogenous participation using maximum simulated likelihood. We allow for the treatment to have an effect on the participation or the sample selection rule and on the main outcome. Applications of this model are frequent in – but no limited to – health economics. We show an application of the model using data from Kenkel and Terza (2001), who investigate the effect of physician advice on the amount of alcohol consumption. Our estimates suggest that in these data (i) neglecting treatment endogeneity leads to a wrongly signed effect of physician advice on drinking intensity, (ii) accounting for treatment endogeneity but neglecting endogenous participation leads to an upward biased estimate of the treatment effect and (iii) advice affects only the drinking intensive margin but not drinking prevalence. Copyright © 2011 John Wiley & Sons, Ltd.

Received 7 October 2010; Revised 8 April 2011; Accepted 5 May 2011

JEL classification: C35; I12; I21

KEY WORDS: count data; drinking; endogenous participation; maximum simulated likelihood; sample selection; treatment effects

1. INTRODUCTION

There are often cases in health economics in which one is interested in the effect of an endogenous dichotomous treatment on an outcome which takes on non-negative integer values with cardinal interpretation (count data). Examples include – but are not limited to – the effect of physician advice on individual alcohol consumption (Kenkel and Terza, 2001), the effect of health status on the number of visits to a general practitioner (Windmeijer and Santos Silva, 1997), or the effect of health insurance coverage on the number of doctor or hospital visits (Riphahn *et al.*, 2003).¹

In all these applications the treatment of interest is likely to be endogenous (*endogenous treatment*).² Receiving physician advice to reduce drinking is certainly not exogenous with respect to the intensity of drinking by patients.³ Similarly, health conditions may not be exogenous with respect to the number of visits to a doctor because individuals who are less concerned with their health are more likely to engage in health damaging behavior and, at the same time, less prone to see a doctor. Also, demand for health insurance is clearly endogenous as high-risk types are expected to buy more comprehensive coverage.

*Correspondence to: Department of Quantitative Social Science, Institute of Education, University of London. 20 Bedford Way, London WC1H 0AL, UK. E-mail: A.Miranda@ioe.ac.uk

¹See Winkelmann (1998) and Greene (2009) for a review of count data models with selectivity.

²This issue is acknowledged and addressed in all the articles we cited.

³In what follows, we will refer to 'drinking' as to alcohol consumption.

In addition to a treatment endogeneity problem, in all these cases one is likely to have also an *endogenous participation* problem.⁴ Indeed, participation into an activity (such as smoking cigarettes, drinking alcohol or seeing a doctor – i.e. the extensive margin) and the intensity of the activity (number of cigarettes/drinks consumed or the number of doctor visits – i.e. the intensive margin) may be two very different processes. For this reason, one might want to allow the two processes to be produced by different data generating processes (DGPs, hereafter). For instance, a person is likely to see a doctor only if she is ill; and the amount of health insurance coverage is likely to have an effect only on the number of times a patient visits the doctor, not on whether she decides to consult a doctor in the first place.

In other cases *sample selection* is the main concern. In a sample of smokers or drinkers, for instance, data on cigarette or alcohol consumption may have not been reported by all individuals, and the data may not be missing at random with respect to the level of alcohol consumption or smoking behavior. For example, heavy smokers or drinkers may have not answered to the survey. In other cases, the researcher may have selected the estimation sample following criteria that are related with individual treatment status or current level of consumption. In all these cases, neglecting sample selection will lead to inconsistent estimates of the treatment effect of interest (e.g. physician advice).

A number of previous papers have suggested strategies for estimating count data models with either sample selection or endogenous treatment, though not both at the same time. Greene (1997), Terza (1998), Winkelmann (1998), Miranda (2004) and Miranda and Rabe-Hesketh (2006) discuss fully parametric methods for estimating count data models based on the Poisson distribution and normally distributed unobserved heterogeneity. Kenkel and Terza (2001) use a flexible Box-Cox specification for the count and normally distributed unobserved heterogeneity to develop a two-step method for estimating the endogenous treatment model. Windmeijer and Santos Silva (1997) discuss a GMM strategy that only requires the specification of the conditional mean of the count y and it is thus less restrictive in terms of the distributional assumptions about y that the researcher needs to impose to achieve a consistent estimator.

The endogenous participation model is closely related to the double-hurdle model of Cragg (1971), the Tobit-type estimator for censored Poisson regression of Terza (1985), the hurdle model of Mullahy (1986), the double-hurdle model of Jones (1989), the two-part model of Mullahy (1998), the endogenous hurdle model of Greene (2009) and the zero-inflated count model of Melkersson and Rooth (2000). All these models are motivated by the idea that individuals must cross one or two hurdles before a strictly positive value of the dependent variable y is observed. Further, the zero outcome is thought to be special in the sense that a large proportion of the individuals in the sample chooses $y = 0$ and that the *participation* decision is qualitatively different from the *intensity of the consumption* decision. For these reasons the models above suggest specifying a different data generating mechanism for zero and strictly positive y . These models have been used to analyze smoking, drinking and fertility behavior among other things. Count data models with endogenous participation are described in Greene (2009) and Mullahy (1998). However, none of the aforementioned models allow for endogenous participation and endogenous treatment at the same time.

To the best of our knowledge, to date only Terza *et al.* (2008) and Li and Trivedi (2009) have suggested strategies to address endogenous treatment and endogenous participation at the same time. Terza *et al.* (2008) put forward a two-step estimator for a *grouped* dependent variable which relies on a joint normality assumption. In this paper, in contrast, we propose an estimation method which is appropriate to deal with endogenous treatment effects and with either sample selection or endogenous participation when the dependent variable is a *count*. Li and Trivedi (2009), on their side, use a Bayesian approach and a two-part model to estimate a model for a *continuous and non-negative* dependent

⁴Here we use a terminology different from that in Greene (2009), who considers self-selection into the treatment as an instance of 'endogenous participation'. In what follows, by 'participation' we generally mean participation in the activity measured by the main (count) outcome variable (e.g. drinking or smoking behavior).

variable with endogenous participation and multivariate treatments. Also in this case, multivariate normality is required. The estimator we propose is similar in spirit to the ones proposed by Terza *et al.* (2008) and Li and Trivedi (2009), relies on the same distributional assumptions and addresses both treatment endogeneity and endogenous participation or both treatment endogeneity and sample selection at the same time.⁵ In addition to the different type of dependent variable that is used in the two aforementioned studies, our approach differs from that of Terza *et al.* in the sense that we use maximum simulated likelihood (MSL). As a consequence, we gain in efficiency with respect to the two-step estimator and obtain correct standard errors in a straightforward fashion (unlike the two-step approach where standard errors of the second step need to be corrected). Our approach differs from Li and Trivedi (2009), as we use a frequentist rather than a Bayesian approach.

We illustrate our estimator using data from Kenkel and Terza (2001), KT hereafter, who study the effect of physician advice on drinking behavior.

The structure of the paper is as follows. In the next section we report a description of the econometric model, focusing on the case of estimating endogenous treatment and endogenous participation. In Section 3 we apply our estimator to KT study on physician advice and drinking behavior. Section 4 summarizes our main findings.

2. THE ECONOMETRIC MODEL

We aim to develop a model for a count variable y_i that is function of a dummy variable T_i representing the i th individual treatment status, with $T_i = 1$ if the individual has been treated and $T_i = 0$ if she has not been treated. The treatment is always observed and, from a theoretical point of view, is a potentially genuine (causal) shifter of the conditional distribution of y_i . We say that T_i is an *endogenous treatment* if treatment status is not random, but there are unobservable individual characteristics affecting T_i that also affect the outcome y_i .

We define a second dummy that represents either a sample selection rule S_i or a participation rule P_i . Although we will refer to models using individual-level data, the individual i subscript is omitted throughout to simplify notation.

In what follows, we describe the main features of the model with endogenous treatment and endogenous participation, while the case with sample selection is illustrated in Appendix A.

2.1. Endogenous participation: $y = 0$ when $P = 0$

The model with (potentially) endogenous participation considers the case where the dependent count variable y for a given individual is *always* zero if the participation dummy P takes on value zero and *can* be positive or null if $P = 1$. In case we aim to investigate the effect of physician advice on drinking behavior, like in KT, we may have three different types of individuals who did not consume alcohol in a given period of time: non-drinkers, drinkers who only drink occasionally and frequent drinkers. The category of non-drinkers may include lifetime abstainers and former drinkers (who are dropped from the analysis by KT), the latter including people who quitted in the period covered by the survey (i.e. recent quitters) and people who are trying to quit. Were the different types of individuals distinguishable (e.g. they declared their type in the survey), a first approach could be to classify occasional and frequent drinkers as participants and non-drinkers as non-participants. In this case the count variable for the number of drinks will include the zeros contributed by occasional or frequent drinkers.⁶ However, if one cannot distinguish among types of zeros – because individuals were asked about their level of alcohol

⁵The model does not address treatment endogeneity, endogenous participation, and sample selection at the same time.

⁶An alternative choice could be to consider only frequent drinkers as participants and occasional drinkers as non-participants together with non-drinkers.

consumption without inquiring whether they are non-drinkers, occasional or frequent drinkers – the natural choice is to pull all zeros together and consider as participants ($P = 1$) only ‘current participants’, i.e. those who consumed a positive number of drinks ($y > 0$) during the period of study. We will use this second approach, as it is more in line with the features of the data used in KT.⁷

The endogenous treatment is denoted as T . The endogenous treatment and the participation dummies are generated according to a continuous latent variable model:

$$T^* = \mathbf{z}'\boldsymbol{\gamma} + v \quad (1)$$

$$P^* = \mathbf{r}'\boldsymbol{\theta} + \varphi T + q \quad (2)$$

where $T = 1(T^* > 0)$, $P = 1(P^* > 0)$, $1(\cdot)$ is the indicator function, and vectors \mathbf{z} and \mathbf{r} represent a set of explanatory variables (including the constant term) with dimension $K_T \times 1$ and $K_P \times 1$, respectively. $\boldsymbol{\gamma}$ and $\boldsymbol{\theta}$ are conformable vectors of coefficients; φ is the coefficient on the treatment in the participation equation; v and q are residual terms. We assume that the count y is generated according to the following conditional cumulative distribution function,

$$G(y|\eta) \equiv \mathbb{P}(y|\eta) = \begin{cases} \text{not defined} & \text{if } P = 0 \\ \mu^y \exp(-\mu)/[1 - \exp(-\mu)]y! & \text{if } P = 1 \end{cases} \quad (3)$$

with

$$y = \begin{cases} 0 & \text{if } P = 0 \\ 1, 2, \dots & \text{if } P = 1 \end{cases} \quad (4)$$

and where $\mathbb{P}(\cdot)$ denotes ‘probability of,’ η is a random variable representing unobserved individual heterogeneity, and $\mu \equiv E[y|\mathbf{x}, T, \eta]$. We use a zero-truncated Poisson distribution for y given $P = 1$. This is done to meet the requirements of the KT data, where it is not possible to distinguish between individuals who did not have any drink because they quitted drinking or are trying to quit, and those who did not have any drink only by chance.⁸ We use a log-linear model for specifying the conditional mean of y given T , P and η :

$$\ln(\mu) = \mathbf{x}'\boldsymbol{\beta} + \delta T + \eta \quad (5)$$

where, again, vector \mathbf{x} represents a $K_y \times 1$ vector of explanatory variables, $\boldsymbol{\beta}$ is a vector of conformable coefficients and δ is the coefficient on the treatment in the equation of the main count response y . Correlation among T , P and y is allowed by imposing some structure on the residuals of Equations (1) and (2),

$$\begin{aligned} v &= \lambda_1 \eta + \zeta \\ q &= \lambda_2 \eta + \xi \end{aligned} \quad (6)$$

where ζ and ξ are ‘idiosyncratic’ error terms and $\boldsymbol{\lambda}' = \{\lambda_1, \lambda_2\} \in \mathbb{R}^2$ are free *factor loadings* to be estimated along the other parameters.

To close the model, we require the covariates to be all exogenous and impose some distributional conditions

$$D(\eta|\mathbf{x}, \mathbf{z}, \mathbf{r}, \zeta, \xi) = D(\eta) \quad (\text{C1})$$

$$D(\zeta|\mathbf{x}, \mathbf{z}, \mathbf{r}, \eta) = D(\zeta|\eta) \quad (\text{C2})$$

$$D(\xi|\mathbf{x}, \mathbf{z}, \mathbf{r}, \eta) = D(\xi|\eta) \quad (\text{C3})$$

$$\zeta \perp \xi | \eta \quad (\text{C4})$$

⁷As we will see below, in KT ‘non-drinkers’ (i.e. individuals who did not drink in the last 12 months, according to their definition), are dropped from the analysis.

⁸Were this piece of information available, a Poisson distribution can be used instead of the zero-truncated Poisson, allowing for two different types of zeros.

where $D(\cdot)$ stands for ‘distribution of’. Condition C1 is the usual random effects assumption, which requires the unobserved individual heterogeneity term η to be independent of all explanatory variables in the system as well as independent of errors ζ and ξ . The conditional independence assumptions in C2 and C3 are weaker than calling for ζ and/or ξ to be independent of the explanatory variables and thus accommodate some limited dependence between control variables and idiosyncratic errors.⁹ C1–C3 together ensure the exogeneity of all explanatory variables \mathbf{x} , \mathbf{z} and \mathbf{r} . Finally, condition C4 requires the idiosyncratic errors to be independent of each other conditional on η . Again, this does not rule out some dependence between ζ and ξ . In what follows, we assume that $\eta \sim N(0, \sigma_\eta^2)$ and that $\zeta|\eta$ and $\xi|\eta$ are both distributed as independent standard normal variates.

The model is identified by restrictions on the covariance matrix and by functional form. So, \mathbf{x} , \mathbf{z} and \mathbf{r} can all have the same elements. However, in spite of formal identification even in the absence of exclusion restrictions, our estimator, like others, may suffer from ‘tenuous identification’ and including equation-specific covariates may be important to ensure the empirical identification of the parameters of interest when real data are used. For this reason, we strongly suggest to use exclusion restrictions whenever it is possible.¹⁰ Note that in our parametrization $\text{Var}(v_i) = (\lambda_1^2 \sigma_\eta^2 + 1)$ and $\text{Var}(q_i) = (\lambda_2^2 \sigma_\eta^2 + 1)$ instead of the usual probit normalization of $\text{Var}(v_i) = \text{Var}(q_i) = 1$. As a consequence, coefficients in (1) and (2) will be larger than the usual probit coefficients. After estimation, one can recover the usual probit parametrization multiplying coefficients in (1) and (2) by a factor of $1/\sqrt{\lambda_1^2 \sigma_\eta^2 + 1}$ and $1/\sqrt{\lambda_2^2 \sigma_\eta^2 + 1}$, respectively.

The use of the Poisson distribution for the analysis of count data has been criticized in the past due to the unattractive feature that conditional mean and conditional variance are constrained to be equal, a property also known as *equidispersion* (see, for instance, Winkelmann, 2008). In contrast, in the present model the introduction of the random term η in the log-linear model for μ allows the count variable y to exhibit *overdispersion* whenever $\sigma_\eta^2 \neq 0$, i.e. if significant unobserved heterogeneity is detected (see Gourieroux, 2000; Miranda and Rabe-Hesketh, 2006). The model, however, cannot be used to analyze count data which exhibit *underdispersion* such as the number of children.¹¹

The correlations between the error terms in y , T^* and S^* are functions of the factor loadings (λ_1, λ_2) and σ_η^2 . In particular, the model implies the following correlations:

$$\rho_{\eta,v} = \frac{\lambda_1 \sigma_\eta^2}{\sqrt{\sigma_\eta^2 (\lambda_1^2 \sigma_\eta^2 + 1)}} \quad (7)$$

$$\rho_{\eta,q} = \frac{\lambda_2 \sigma_\eta^2}{\sqrt{\sigma_\eta^2 (\lambda_2^2 \sigma_\eta^2 + 1)}} \quad (8)$$

$$\rho_{v,q} = \frac{\lambda_1 \lambda_2 \sigma_\eta^2}{\sqrt{(\lambda_1^2 \sigma_\eta^2 + 1)(\lambda_2^2 \sigma_\eta^2 + 1)}} \quad (9)$$

⁹If claiming independence between all explanatory variables and the unobserved heterogeneity term η is judged untenable for a particular application, instead of requiring condition C1 one could follow Mundlak (1978) and Chamberlain (1980) approach and assume $\eta|\mathbf{w} \sim N(\mathbf{w}'\boldsymbol{\psi}, \sigma_\eta^2)$, for a vector \mathbf{w} that can contain some elements of \mathbf{x} , \mathbf{z} , and \mathbf{r} and where $\boldsymbol{\psi}$ is a vector of conformable coefficients. This assumption imposes some restrictions to the way explanatory variables and the unobserved heterogeneity term η can be related (namely, linearity and some cross-equation coefficient restrictions) but allows at least some dependence.

¹⁰Other estimators are formally identified without exclusion restrictions, but in empirical work it is common practice to include them anyway to help identification. See, among others, KT for endogenous treatment effects applied to count data, Puhani (2000) for the Heckman selection model, Monfardini and Radice (2008) for the bivariate probit model and Keane (1992) for the multinomial probit model.

¹¹To the knowledge of the authors no method has been suggested in the literature that could deal with underdispersed count data and either a sample selection or an endogenous treatment effect, let alone the two problems together.

The treatment T is an exogenous variable in the main response equation whenever $\rho_{\eta,v} = 0$. Similarly, if $\rho_{\eta,q} = 0$ participation is exogenous in the main response equation. And, if $\rho_{\eta,v} = \rho_{\eta,q} = 0$ one can obtain consistent estimates of δ on the basis of a simple Poisson regression fitted on the sub-sample for which $y > 0$ (i.e. ‘participants’).

Let $\mathbb{P}_P(0|\eta)$ denote the conditional probability of $P = 0$ given η and $\mathbb{P}_P(1|\eta)$ the conditional probability of $P = 1$ given η . Here, to simplify notation, we do not explicitly write the conditioning on observable variables. In a similar fashion, $\mathbb{P}_T(\tau|\eta)$ represents the probability of $T = \tau$ given η , with $\tau = \{0, 1\}$. The log-likelihood function is then

$$\begin{aligned} \log(L) = & \sum_{i, P_i=0} \sum_{\tau} \omega_{\tau} \ln \left\{ \int \mathbb{P}_P(0|\eta) \mathbb{P}_T(\tau|\eta) \phi(\eta) d\eta \right\} \\ & + \sum_{i, P_i=1} \sum_{\tau} \omega_{\tau} \ln \left\{ \int \mathbb{P}_P(1|\eta) \mathbb{P}_T(\tau|\eta) G(y|\eta) \phi(\eta) d\eta \right\} \end{aligned} \quad (10)$$

where $\phi(\cdot)$ is the density of a normal variate with mean zero and variance σ_{η}^2 , $\omega_0 = 1(T = 0)$ and $\omega_1 = 1(T = 1)$. We will refer to this model as the Endogenous Participation Endogenous Treatment (EPET-) Poisson model.

The integrals in Equation (10) do not have a closed form solution and must be numerically evaluated. We use MSL (for a detailed discussion on MSL, see, Train, 2003). To evaluate the integrals we use Halton sequences instead of uniform pseudorandom sequences. Halton draws have been shown to achieve high precision with fewer draws than uniform pseudorandom sequences because they have a better coverage of the $[0, 1]$ interval. A modified Newton–Raphson algorithm is used for maximization, using analytical first derivatives and numerical second derivatives. At convergence Eicker–Huber–White robust standard errors are computed.

The use of a common latent factor structure like the one written in (6) has four main advantages over the alternative of specifying a multivariate normal distribution for v , q and η (see Deb and Trivedi, 2006). First, the common latent variable approach can be used quite flexibly to combine appropriately chosen conditional and marginal distributions that generate the joint distribution that the researcher wants to use. Second, latent factors have a natural interpretation as proxies for unobserved covariates since they enter into the equations in the same way as observed covariates. The factor loadings can therefore be interpreted in much the same way as coefficients on observed covariates can. Third, it provides a parsimonious representation of error correlations in models with a large number of equations. Related to this, and quite importantly for computational feasibility, the latent variable approach transforms a problem in which calculation of the log-likelihood involves the computation of a three-dimensional integral into a problem where only a one-dimensional integral needs to be computed.

The cost of using the one (latent) factor structure is that it implies some variance–covariance restrictions.¹² Carneiro *et al.* (2003) note that these restrictions can have an arbitrary content and conclude that it is important to appeal to economic theory to justify any specific identification scheme. In our case, there are three implicit variance–covariance restrictions as the six elements of the variance–covariance matrix must be retrieved from only three parameters (λ_1 , λ_2 and σ_{η}^2).¹³ Hence, the model we propose is especially useful in cases in which the endogeneity of both treatment and

¹²Latent factor models are very popular in health economics. For some recent examples of normally distributed one-latent factor models see KT, Terza *et al.* (2008) and Terza (2009).

¹³Let us define the variance–covariance matrix of the error terms as:

$$V = \begin{bmatrix} \sigma_{\eta}^2 & \sigma_{\eta,v} & \sigma_{\eta,q} \\ & \sigma_v^2 & \sigma_{v,q} \\ & & \sigma_q^2 \end{bmatrix} \quad (11)$$

The three restrictions are (i) $\sigma_v^2 = \lambda_1^2 \sigma_{\eta}^2 + 1$, (ii) $\sigma_q^2 = \lambda_2^2 \sigma_{\eta}^2 + 1$, and (iii) $\sigma_{v,q} = (\sigma_{\eta,v} \cdot \sigma_{\eta,q}) / \sqrt{\sigma_v^2 \sigma_q^2}$ (or $\rho_{v,q} = \rho_{\eta,v} \cdot \rho_{\eta,q}$).

participation are likely to be determined by a common unobservable variable entering all error terms. In the case of drinking advice, the unobservable factor may be an individual's intertemporal discount rate. Individuals with a high intertemporal discount rate are more likely to engage in health damaging activities such as alcohol consumption, to be in bad health, and to receive drinking advice when seeing a doctor. In particular, they are both more likely to drink and to drink a high number of drinks. To put it simply, the omission of a single unobservable which enters all three equations is likely to generate a positive correlation between all three error terms. The model is also suitable to cases in which two or more highly correlated unobservable variables enter all equations. In such a case using a one latent factor model approach may be a good strategy because after conditioning on a single *summary* latent factor η , the remaining components of the error terms in the three equations can be expected to be approximately orthogonal (see condition C4). In general, it is the responsibility of each applied researcher to judge whether the estimator we propose here is suitable to his/her specific problem.

In Appendix B we investigate the finite sample performance of the EPET-Poisson model and its alternatives in a Monte Carlo simulation study where the true DGP is EPET-Poisson. We compare: (i) EPET-Poisson, (ii) a model in which only the endogeneity of the treatment is addressed while participation is neglected (the ET-Poisson model) and (iii) a model in which the endogenous participation is addressed while treatment is taken to be exogenous (EP-Poisson). Results show that EPET-Poisson is consistent and asymptotically normally distributed. Further, we find evidence showing that inference can be seriously misleading if one fits ET-Poisson or EP-Poisson to data that are truly EPET-Poisson.

3. AN APPLICATION TO THE EFFECT OF PHYSICIAN ADVICE ON DRINKING BEHAVIOR

In this section, we apply the EPET-Poisson model to the problem of estimating the treatment effect of physician advice on alcohol consumption using data from KT.

3.1. The KT's study

For the purpose of illustration, we use the same data – the 1990 National Health Interview Survey (NHIS) – and adopt the same empirical specification and exclusion restrictions used by KT.¹⁴ Our aim is to show how the estimates of treatment effects are sensitive to various assumptions about the endogeneity of treatment status and drinking participation. In the KT study drinking is measured as the number of alcoholic drinks consumed in the last 2 weeks.¹⁵ Physician advice about drinking is built from respondents' answers to the following question: 'Have you ever been told by a physician to drink less?'¹⁶

¹⁴Although the model is formally identified without exclusion restrictions, KT provide an additional source of identification 'through exclusion restrictions involving a set of 11 variables related to health insurance status, physician contacts and health problems' (p. 176). KT use an overidentifying test to check the validity of their exclusion restrictions. In particular, they assume that the insurance-related 'instruments' (MEDICARE, MEDICAID, CHAMPUS and HLTHINS) are valid and test for the validity of the remaining 'instruments'. The Wald test supports their choice (see footnote 13, p. 176). Hence, in our paper, we assume that these exclusion restrictions are valid and report results using their empirical specification. The data are available at <http://www.econ.queensu.ca/jae/2001-v16.2/kenkel-terza/>.

¹⁵KT state 'This is calculated as the product of self-reported drinking frequency (the number of days in the past 2 weeks with any drinking) and drinking intensity (the average number of drinks on a day with any drinking)', (p. 171–172).

¹⁶A potential weakness of these data is that the timing of the physician advice is likely to differ among individuals. One could expect that the further in the past the advice was given, the less influence the advice has on contemporaneous drinking behavior. In such a case, one would obtain an under-estimate of the effect of drinking advice in our application which considers 'lifetime physician advice' compared with studies considering current or recent physician advice. Moreover, some individuals may have already quitted drinking because of the physician's advice, introducing a potential sample selection bias in the estimates since KT exclude former drinkers from their analysis. For these reasons, the specific estimates of treatment effects we present in this paper must be interpreted with caution.

The authors drop from the analysis lifetime abstainers and former drinkers, i.e. individuals who used to drink but with no drinking in the past year. Because the physician advice to cut drinking was given as a way of reducing high blood pressure, they focus only on men who have drunk alcohol at least once in the last 12 months ('drinkers') and report having been told at some time that they had high blood pressure. KT observe in their sample that 21% of drinkers (according to their definition) did not drink at all in the last 2 weeks. Various reasons may be behind the excess of zeros. First, it could be that zeros are contributed by recent quitters (less than one year) or people who were actively trying to quit drinking in the last 12 months. We define these individuals as 'quitters'. Second, $y = 0$ could also be contributed by individuals who drink only in very special occasions such as weddings or parties (occasional drinkers). Moreover, $y = 0$ could be contributed by frequent drinkers who, by chance, did not drink any alcohol in the past 2 weeks. However, this last scenario is less likely as 2 weeks are a period long enough to expect a strictly positive number of drinks to be consumed by frequent drinkers. Unfortunately, it is not possible to know whether individuals classify themselves as 'quitters', occasional or frequent drinkers. As a consequence, we cannot distinguish between different types of zeros.

Clearly, the fact that a good proportion (perhaps the majority) of these zeros are likely to be contributed by occasional drinkers and 'quitters' suggests that the excess zeros cannot be ignored. The authors acknowledge this and account for the excess zeros by using a flexible functional form for the conditional mean of drinking based on the inverse Box-Cox transformation. An alternative way of addressing this issue, which we follow here, is to treat the zeros and the positive drinking outcomes as if they were generated by two separate DGPs (see Terza, 1998). More details on the data and the covariates used are available in the original study. Table I reports the definitions and the means of all the variables, which match the corresponding means in KT. The table also provides information on which variables are used to help the identification of the model (the exclusion restrictions).

3.2. Results from EPET-Poisson

3.2.1. The effect of the treatment. In this section, we focus only on the effect of the treatment of interest (physician advice). As we said, we use the same specifications (and exclusion restrictions) as the original article for the treatment and the drinking intensity equations.

The first column of Table II reports the marginal effects from a Poisson model where the potential endogeneity of the treatment – physician advice – is not addressed.¹⁷ The results are similar to those reported by KT in the models where physician advice is considered exogenous (see Table III in their article): advice appears to have a counterintuitive *positive* effect on drinking that is statistically significant at 1%. Column (2) reports the marginal effects of physician advice on the probability of drinking obtained from a simple probit model, and also in this case, advice turns out to be positively associated with drinking. Column (3) reports the marginal effects when the potential endogeneity of advice is taken into account but endogenous participation is neglected using the ET-Poisson model. This model assumes that both zeros and positive y outcomes are produced by the same DGP but accounts for the endogeneity of the treatment.¹⁸ This model (and the following models employing MSL) was estimated using 1600 Halton draws.¹⁹ First, note that the correlation between the errors in the drinking intensity and the physician advice equation $\rho_{\eta,v}$ is positive, as expected, and statistically significant at 1%. Hence, advice is endogenous with respect to the level of drinking. In other words, individuals who have a higher latent propensity to drink are also more likely to receive advice. Second,

¹⁷In analogy to KT, marginal effects are evaluated at the *median* value of the dependent variable.

¹⁸More details are available upon request from the authors. The model consists of two equations, the drinking equation in which both zero and positive consumption are modelled through a Poisson model with unobserved heterogeneity and a physician advice equation. The model is estimated using MSL, using a one latent factor structure and assuming normality of the error terms.

¹⁹Using 2000 Halton draws produced only negligible changes in coefficients and standard errors of the estimates.

Table I. Variable definitions and descriptive statistics

Variable name	Definition	Mean	SD
<i>Dependent variable</i>			
<i>y</i>	Total drinks last 2 weeks	14.697	22.753
<i>Treatment variable</i>			
<i>T</i>	Physician advice to reduce drinking	0.278	—
<i>Control variables^a</i>			
EDITINC	Income (\$1000)	2.575	5.008
AGE30	30 < age ≤ 40	0.180	—
AGE40	40 < age ≤ 50	0.195	—
AGE50	50 < age ≤ 60	0.182	—
AGE60	60 < age ≤ 70	0.199	—
AGEGT70	Age > 70	0.122	—
EDUC	Years of schooling	12.925	3.087
BLACK	Black	0.133	—
OTHER	Non-white	0.018	—
MARRIED	Married	0.645	—
WIDOW	Widowed	0.052	—
DIVSEP	Divorced or separated	0.160	—
EMPLOYED	Employed	0.666	—
UNEMPLOY	Unemployed	0.029	—
NORTHE	Northeast	0.217	—
MIDWEST	Midwest	0.275	—
SOUTH	South	0.295	—
<i>Excluded variables^b</i>			
MEDICARE	Insurance through Medicare	0.252	—
MEDICAID	Insurance through Medicaid	0.031	—
CHAMPUS	Military insurance	0.059	—
HLTHINS	Health insurance	0.815	—
REGMED	Regular source of care	0.821	—
DRI	See same doctor	0.721	—
MAIORLIM	Limits on major daily activity	0.086	—
SOMELIM	Limits on some daily activity	0.077	—
HVDIAB	Have diabetes	0.061	—
HHRTCOND	Have heart condition	0.146	—
HADSTROKE	Had stroke	0.036	—

^aThese are the variables included in both the main equation (and the endogenous participation equation) and the endogenous treatment equation.

^bThese are the variables included only in the endogenous treatment equation.

Note: This table reports the definitions, the means and the standard deviations (SD) for the variables used in KT. Data refer to the 1990 NHIS. The estimation sample includes 2467 observations.

the marginal effect of physician advice turns out to be negative, statistically significant, and amounts to a bit less than $-5\frac{1}{2}(-5.4)$ drinks per 2 weeks. Both results suggest that the positive effect of *T* on *y* that is shown by the Poisson model with exogenous treatment is spuriously driven by a positive bias which results from the fact that individuals endogenously sort themselves into the treatment. In other words, those receiving advice were also the heaviest drinkers.

Column (5) reports the effects of physician advice estimated from the EP-Poisson model in which endogenous participation is addressed while treatment status is exogenous. The effect on both drinking participation and drinking intensity turn out to be statistically significant and positive, like in the simple Probit and simple Poisson models, respectively. The correlation between the errors in the two equations is positive but statistically insignificant.

When the intensive margin (i.e. drinking participation) and the extensive margin (i.e. number of drinks conditional on strictly positive drinking) are allowed to be generated by different DGPs and physician advice to be endogenous with the EPET-Poisson model, in column (7) the effect of physician advice falls in absolute value by more than one drink per 2 weeks, to -4.1 (-25% in absolute value),

Table II. Marginal effects of physician advice on the number of drinks consumed in the last 2 weeks and on the probability of drinking

	Poisson	Probit	ET-Poisson	EP-Poisson		EPET-Poisson	
	y^a (1)	$\Pr(y > 0)^b$ (2)	y^a (3)	$\Pr(y > 0)^b$ (4)	$y > 0^a$ (5)	$\Pr(y > 0)^b$ (6)	$y > 0^a$ (7)
Physician advice (T)	3.679*** (0.558)	0.079*** (0.017)	-5.395*** (0.386) 0.832*** (0.029)	0.079*** (0.017)	3.377*** (0.557)	-0.045 (0.049)	-4.072*** (0.864)
$\hat{\rho}_{\eta,v}$						0.689*** (0.092)	
$\hat{\rho}_{\eta,q}$				0.082 (0.141)		0.378*** (0.088)	
$\hat{\rho}_{v,q}$						0.261*** (0.084)	
$\hat{\sigma}_\eta^2$			2.190*** (0.069)	1.207*** (0.023)		1.456*** (0.099)	
N.obs.	2467	2467	2467	2467	2467	2467	2467
Log-likelihood	-32 263	-1247	-10 184	-8660		-10 062	
$\Pr(\widehat{y} = 0)^c$	0.00	0.21	0.11	0.21		0.22	
BIC	70 360		23 418	20 675		20 671	

***Significant at 1%. Eicker–Huber–White robust standard errors in parentheses.

Note: y is the number of alcoholic drinks consumed in the last two weeks, $\Pr(y > 0)$ is the probability of drinking in the last 2 weeks, $y > 0$ is the number of alcoholic drinks consumed in the last 2 weeks conditional on drinking and T is a dichotomous indicator of individual treatment status. Estimation refers to the 1990 NHIS with the sample selection and covariates used in KT. ET, EP and EPET stand for Endogenous Treatment, Endogenous Participation and Endogenous Participation Endogenous Treatment, respectively. ET-Poisson, EP-Poisson and EPET-Poisson models were estimated using MSL and 1600 Halton draws. The joint Wald test statistic for $\rho_{\eta,v} = \rho_{\eta,q} = \rho_{v,q} = 0$ in the EPET-Poisson model, distributed as a $\chi^2(3)$, is 57.904 (p -value = 0.00).

^aThe marginal effect is computed for a discrete change in T (from 0 to 1) at the sample median of the dependent variable, in analogy to KT.

^bThe marginal effect is computed for a discrete change in T (from 0 to 1) at the sample mean of the other independent variables.

^cProbability of the zero outcome predicted by the model.

^dBayesian information criterion. For the sake of comparability all BICs refer to a three-equation model. In the ET-Poisson, the BIC refers to the ET-Poisson and the probit equation for exogenous participation; in the EP-Poisson, the BIC refers to the EP-Poisson and the probit equation for exogenous treatment. The BIC in columns (1)–(2) refers to the probit models for exogenous participation and exogenous treatment and the Poisson model with exogenous participation and exogenous treatment (including the zeros).

and remains highly statistically significant.²⁰ Physician advice turns out to be endogenous with respect to drinking participation and $\rho_{v,q}$ is positive, which is consistent with the correlation found between advice and drinking intensity (i.e. $\rho_{\eta,v} > 0$) and with our theoretical predictions at the end of Section 2.1. It is also important to notice that the EPET-Poisson model shows that physician advice has no effect on the likelihood of drinking alcohol (column (6)). This result is in sharp contrast with the one obtained from the simple probit. So, our findings suggest that the positive association between drinking and physician advice found in the probit model was spuriously driven by unobserved heterogeneity.

Comparison of the ET-Poisson and the EPET-Poisson models allows a better understanding of the effect of physician advice, which does not seem to induce people not to drink but simply to cut their two-week drinking. An explanation of this could be that the population is composed by a mixture of two types: light drinkers and heavy drinkers. Light drinkers do not quit, following physician advice, because they do not believe that their limited drinking is damaging their health. Heavy drinkers, in a similar fashion, do not quit simply because they have higher levels of addiction. As a consequence,

²⁰We included the same set of controls both in the drinking participation and in the drinking intensity equations. The same is done in Terza *et al.* (2008). In general, unlike in sample selection models in which there might be specific factors affecting selection (e.g. survey attrition) but not necessarily the intensity of consumption, in the case of endogenous participation it is hard to think of variables affecting either the intensive or the extensive margin only.

Table III. Marginal effects of covariates on the physician advice equation

	ET-Poisson	EPET-Poisson
EDITINC	0.000 (0.002)	0.000*** (0.000)
AGE30	0.078** (0.035)	0.098*** (0.035)
AGE40	0.041 (0.035)	0.055 (0.035)
AGE50	0.023 (0.034)	0.039 (0.036)
AGE60	0.022 (0.040)	0.047 (0.041)
AGEGT70	0.041 (0.050)	0.057 (0.051)
EDUC	−0.009*** (0.003)	−0.010*** (0.000)
BLACK	0.108*** (0.029)	0.106*** (0.026)
OTHER	0.089 (0.063)	0.079 (0.063)
MARRIED	0.047* (0.026)	0.037 (0.026)
WIDOW	0.100** (0.049)	0.083* (0.049)
DIVSEP	0.102*** (0.034)	0.080** (0.033)
EMPLOYED	−0.006 (0.027)	−0.013 (0.027)
UNEMPLOY	0.084 (0.055)	0.042 (0.049)
NORTHE	0.023 (0.025)	0.030 (0.026)
MIDWEST	−0.021 (0.024)	−0.009 (0.024)
SOUTH	−0.017 (0.024)	−0.006 (0.024)
<i>Excluded variables^a</i>		
MEDICARE	0.014 (0.045)	0.011 (0.043)
MEDICAID	0.006 (0.033)	0.016 (0.037)
CHAMPUS	−0.049** (0.022)	−0.069*** (0.024)
HLTHINS	0.041* (0.025)	0.065** (0.026)
REGMED	0.011 (0.021)	−0.006 (0.025)
DRI	0.051 (0.034)	0.008 (0.031)
MAIORLIM	0.011 (0.027)	0.000 (0.028)
SOMELIM	0.108*** (0.034)	0.085** (0.034)
HVDIAB	0.063*** (0.023)	0.068*** (0.023)
HHRTCOND	0.029 (0.036)	0.059 (0.040)
HADSTROKE	−0.009 (0.029)	−0.020 (0.030)
<i>F-test excluded variables</i>	49.22	53.8
	[0.00]	[0.00]

***Significant at 1%; **significant at 5%; *significant at 10%. Eicker–Huber–White robust standard errors in parentheses, *p*-values in brackets.

Note: The dependent variable is the probability of receiving physician advice of not drinking. The table reports for each variable the marginal effect computed at the sample mean (of the other covariates) and the *F*-tests for the ‘exclusion restrictions’ in the ET- and the EPET-Poisson models.

^aVariables excluded from the drinking equation to help identification.

physician advice may have an effect only on the drinking intensive margin but not on drinking prevalence.²¹

Table II shows the presence of significant unobserved heterogeneity in the three choices (as σ_{η}^2 is statistically different from zero). The last two lines of the table also clearly suggest that the EPET-Poisson model is obviously much better at predicting the zeros and that it fits the data better than the alternative models (Poisson, EP-Poisson and ET-Poisson) as indicated by the lowest value of the Bayesian information criterion.²²

Two notes are worth of mentioning. First, in KT’s specific case pooling the intensive and the extensive margins and forcing the DGP to be the same for the two choices are not very harmful because

²¹We estimated the EPET-Poisson model also without exclusion restrictions. In this case, all the correlations between the error terms turned out to be positive but statistically insignificant. The effect of physician advice was instead significant in the drinking participation equation (marginal effect = 0.07, SE = 0.03) but not in the drinking intensity equation (marginal effect = −1.35, SE = 3.05). Hence, our application shows that when the same variables are included in all equations the data may not be able to identify the variance and covariance parameters, and confirms that exclusion restrictions are useful to help the empirical identification of the parameters of interest. These additional results are available from the corresponding author upon request.

²²The EP-Poisson is nested into the EPET-Poisson model while the EPET-Poisson (or the EP-Poisson) and the ET-Poisson are non-nested, as in the latter the count variable also includes zeros.

the effect of the treatment on the two outcomes goes in the same direction, although our estimates suggest that only the intensive margin is significantly affected by physician advice. Furthermore, the correlations between unobservables in the endogenous treatment and drinking intensity, and between the unobservables in the endogenous treatment and the endogenous participation, all have the same sign. Second, KT use a flexible functional form (the non-linear inverse Box-Cox form). Although imposing the same DGP for the intensive and extensive drinking margins, the modeling strategy followed by KT produces a marginal effect of the treatment (about -4.5 drinks) that is somewhat between the one reported by a model that only deals with the endogenous treatment and the one obtained from a model that deals with both endogenous treatment and endogenous participation (the EPET-Poisson model). Clearly, in other applications the consequences of neglecting endogenous participation may be more substantial.

In order to have an idea of the goodness of the exclusion restrictions, Table III reports the marginal effects for the physician advice equation and Wald tests for the exclusion of the variables identifying the model over and above functional form and covariance restrictions. In both the ET- and the EPET-Poisson models Wald tests suggest that the ‘excluded variables’ are highly statistically significant and the model is unlikely to suffer from weak identification.

3.2.2. The effect of other covariates. The main advantage of a model not imposing the same DGPs on the intensive and the extensive drinking margins is that not only the effect of the treatment but also that of other covariates are allowed to differ across the two choices. Think of the effect of parental supervision, for instance, or strictness of parenting styles, on youngsters’ smoking behavior. In this case, parenting style is likely to affect the likelihood of smoking participation but it is rather unlikely to affect the quantity of cigarettes smoked given participation. Similarly, alcohol (cigarette) taxes and prices are more likely to affect the quantity of drinks (cigarettes) consumed than the drinking (smoking) prevalence itself. Table IV reports the marginal effects (at the sample mean) of the other covariates estimated in the EPET-Poisson model and shows which is the relevant margin (intensive, extensive or both) affected by the regressors. Just to take a few examples, it is interesting to notice that years of education are positively associated with the probability of drinking but negatively associated with the average number of drinks consumed. Individuals in their forties and fifties drink less on average, but this effect is entirely accounted for by their lower probability of drinking. We also reported the marginal effects obtained from ET-Poisson and EP-Poisson models for the sake of completeness.

3.3. Results of the SSET-Poisson model

As we already said, KT exclude from the estimation sample lifetime abstainers and former drinkers (those who used to drink but did not consume any alcoholic drink during the past year). Although the exclusion of the former is justified by the specific application, as for them physician advice cannot have a causal effect on drinking, the exclusion of the latter may cause a sample selection bias in the analysis. For instance, people who are more sensitive to physician advice may have already quitted drinking, with the consequence of finding in the estimation sample only heavy drinkers for whom the effect of the treatment may be lower than in the general population of drinkers.²³

Hence, for illustrative purposes we show here how the Sample Selection Endogenous Treatment (SSET-) Poisson model could be useful to address this potential sample selection issue. Details on the SSET-Poisson model are reported in Appendix A.²⁴

²³Another potential problem with this sample selection is that the decision to quit drinking may be affected by health status, so that in the sample one is likely to observe only ‘healthy’ drinkers. KT argue that this is likely to induce only a small selection bias as in the NHIS only 12% of individuals declare not to drink because of health problems.

²⁴The treatment equation includes all controls while the sample selection and the drinking equations do not include the ‘excluded variables’ in Table I. The SSET-Poisson has not only an advantage over the EPET-Poisson model since it addresses the sample

Table IV. Marginal effects of other covariates

	ET-Poisson	EP-Poisson		EPET-Poisson	
	y	$\Pr(y > 0)$	$y > 0$	$\Pr(y > 0)$	$y > 0$
EDITINC	0.047 (0.040)	0.002 (0.002)	0.091*** (0.027)	0.001** (0.001)	0.074** (0.030)
AGE30	0.961 (0.735)	-0.041 (0.035)	-0.212 (0.892)	-0.027 (0.015)	1.298 (0.869)
AGE40	-0.158 (0.683)	-0.063* (0.036)	-1.013 (0.913)	-0.049*** (0.016)	0.101 (0.833)
AGE50	-0.991* (0.520)	-0.077** (0.037)	-1.903*** (0.951)	-0.063*** (0.018)	-1.071 (0.839)
AGE60	-0.763 (0.646)	-0.037 (0.038)	-2.394** (1.068)	-0.025 (0.018)	-1.396* (0.771)
AGEGT70	-1.192* (0.651)	-0.008 (0.042)	-3.081*** (0.967)	0.001 (0.020)	-2.140*** (0.739)
EDUC	-0.101 (0.067)	0.010*** (0.003)	-0.221** (0.091)	0.008*** (0.002)	-0.324*** (0.086)
BLACK	0.298 (0.663)	-0.001 (0.025)	-2.695*** (0.693)	0.014 (0.013)	-0.93 (0.600)
OTHER	-1.068 (0.891)	0.030 (0.062)	-4.515*** (1.169)	0.042 (0.026)	-3.215*** (1.093)
MARRIED	0.024 (0.541)	-0.008 (0.026)	-0.941 (0.810)	-0.006 (0.012)	-0.500 (0.609)
WIDOW	2.396** (1.087)	0.001 (0.046)	-0.072 (1.419)	0.011 (0.022)	1.308 (1.482)
DIVSEP	2.061** (0.848)	0.002 (0.032)	1.557 (1.153)	0.009 (0.015)	2.103** (0.978)
EMPLOYED	0.213 (0.538)	0.077*** (0.025)	-0.297 (0.830)	0.072*** (0.013)	-0.464 (0.711)
UNEMPLOY	5.240*** (1.972)	0.091** (0.039)	3.234 (4.260)	0.092*** (0.017)	3.286* (1.887)
NORTHE	-0.192 (0.516)	-0.038 (0.027)	-0.266 (0.775)	-0.034*** (0.013)	0.060 (0.783)
MIDWEST	-1.284*** (0.474)	-0.046* (0.026)	-0.897 (0.751)	-0.046*** (0.012)	-0.965 (0.703)
SOUTH	-1.103** (0.468)	-0.044* (0.026)	-0.914 (0.738)	-0.043*** (0.012)	-0.864 (0.740)

***Significant at 1%; **significant at 5%; *significant at 10%. Eicker–Huber–White robust standard errors in parentheses.

Note: y is the number of alcoholic drinks consumed in the last 2 weeks, $\Pr(y > 0)$ is the probability of drinking in the last 2 weeks and $y > 0$ is the number of alcoholic drinks consumed in the last 2 weeks conditional on drinking. For each variable marginal effects are computed at the sample means of the other independent variables.

In Table V we report results from simple Poisson, probit for sample selection, ET-Poisson, Sample Selection (SS-) Poisson and SSET-Poisson models.²⁵ The probit model for sample selection shows that advice does not appear to have an effect on the sample selection, i.e. the probability of not being a former drinker. The SS-Poisson model, in which only the sample selection issue is addressed while the treatment is considered as exogenous, shows the existence of a severe sample selection issue. In particular, drinkers appear to be negatively selected. This is, from an economic point of view, counter intuitive.²⁶ However, it can be shown that neglecting treatment endogeneity produces this odd and spurious result.²⁷ The effect of the treatment on the sample selection is statistically insignificant like in the simple probit model, whereas the effect on current drinking is positive and statistically significant like in the simple Poisson model. However, the SSET-Poisson model shows that this last result is completely driven by unobserved heterogeneity, namely by the positive correlation between η and v . When the endogeneity of physician advice is addressed, and the model accounts for the fact that heavy drinkers are also more likely to receive advice, the sample selection issue disappears and the estimate of the treatment effect becomes very close to that obtained from the ET-Poisson. The result of an insignificant effect of physician advice on the probability of being a former drinker is consistent with the finding of an insignificant effect of physician advice on drinking participation for current drinkers obtained in the previous section with the EPET-Poisson model.

Simple Poisson, ET-Poisson and SS-Poisson are all nested within the SSET-Poisson model. Hence, model selection can be based on significance tests for the correlation coefficients.

selection issue but also a disadvantage since it forces the drinking participation and the drinking intensity decisions to be produced by the same DGP.

²⁵As the data released by KT do not include former drinkers, we had to resort to the original 1990 NHIS data. The data used in this section are available in Stata format from the authors along with complete details on how the data set was created.

²⁶We expect heavy drinkers not to have quitted in the past.

²⁷If there is severe treatment endogeneity (e.g. $\rho_{v,q}$ and $\rho_{\eta,v}$ are positive and high) and sample selection is either small or absent in the true DGP (SSET-Poisson or EPET-Poisson), statistical theory predicts that SS-Poisson should find – false – evidence of sample selection. To see why, notice that in a trivariate normal distribution with such correlation matrix the partial correlation between S^* and y , net of the effect of T^* , will be negative (see for instance, Cox and Wermuth, 1996, p. 69).

Table V. Marginal effects of physician advice on the number of drinks consumed in the last 2 weeks and on the probability of sample selection – sample selection model

	Poisson	Probit	ET-Poisson	SS-Poisson		SSET-Poisson	
	y^a (1)	$\Pr(S=1)^b$ (2)	y^a (3)	$\Pr(S=1)^b$ (4)	y^a (5)	$\Pr(S=1)^b$ (6)	y^a (7)
Physician advice (T)	3.679*** (0.558)	0.001 (0.016)	−5.395*** (0.386)	−0.014 (0.014)	4.932*** (1.052)	−0.111 (0.145)	−5.465*** (0.809)
$\hat{\rho}_{\eta,v}$			0.832*** (0.029)			0.855*** (0.054)	
$\hat{\rho}_{\eta,q}$				−0.986*** (0.002)		0.199 (0.245)	
$\hat{\rho}_{v,q}$						0.207 (0.229)	
$\hat{\sigma}_\eta^2$			2.190*** (0.069)	3.304*** (0.005)		2.292*** (1.195)	
N.obs.	2467	4952	2467	4952		4952	
Log-likelihood	−32 263	−3432	−10 184	−12 132		−15 066	
$\Pr(\widehat{y} = 0)^c$	0.00	—	0.11	0.00		0.28	

***Significant at 1%. Eicker–Huber–White robust standard errors in parentheses.

Note: y is the number of alcoholic drinks consumed in the last two weeks, S is an indicator for sample selection, $\Pr(S=1)$ is the probability of sample selection (not being a former drinker) and T is a dichotomous indicator of individual treatment status. Estimation refers to the 1990 NHIS with the sample selection and covariates used in KT, but including former drinkers. ET, SS and SSET stand for Endogenous Treatment, Sample Selection and Sample Selection Endogenous Treatment, respectively. ET-Poisson, SS-Poisson and SSET-Poisson models were estimated using MSL and 1600 Halton draws. The joint Wald test statistic for $\rho_{\eta,v} = \rho_{\eta,q} = \rho_{v,q} = 0$ in the SSET-Poisson model, distributed as a $\chi^2(3)$, is 1592 (p -value = 0.00).

^aThe marginal effect is computed for a discrete change in T (from 0 to 1) at the sample median of the dependent variable, in analogy to KT.

^bThe marginal effect is computed for a discrete change in T (from 0 to 1) at the sample mean of the other independent variables.

^cProbability of the zero outcome predicted by the model.

4. CONCLUDING REMARKS

In this paper, we have proposed a Full Information Maximum Likelihood estimator for count data models with endogenous treatment effects and either sample selection or endogenous participation, which is implemented using MSL. Sample selection occurs, for instance, when the main outcome is missing for some individuals – because of item or survey non-response – and the data are not missing at random, or when the researcher selected the estimation sample following criteria that could be related to individual treatment status or the outcome of interest. In contrast, endogenous participation occurs when participation into an activity (e.g. smoking or drinking behavior) and the intensity of the activity are produced by two different, but correlated, DGPs.

For illustrative purposes, we have applied our proposed estimator to KT data on physician advice and drinking. Our estimates suggest that in these data (i) neglecting treatment endogeneity leads to a perversely signed effect of physician advice on drinking intensity, (ii) accounting for treatment endogeneity but neglecting endogenous participation leads to an upward biased estimate of the treatment effect of physician advice on drinking intensity and (iii) advice only affects the drinking intensive margin but not drinking prevalence.

Since KT dropped from the sample former drinkers, we have also assessed the sensitivity of treatment effects estimates to this sample selection. Our results suggest that in the 1990 NHIS data physician advice does not seem to affect the likelihood of being a former drinker, and estimates of treatment effects are not biased by restricting the analysis only to current drinkers.

Two main limitations of our estimator are that (i) it does not allow to address both endogenous participation and sample selection at the same time, (ii) it uses a one latent factor structure and it would be interesting for future research to remove both of these restrictive features and extend the range of problems to which it can be fruitfully applied.

APPENDIX A: ENDOGENOUS SAMPLE SELECTION: Y MISSING WHEN $S = 0$

In some cases, the most pressing data problem may be that a non-negligible proportion of the interviewed individuals did not respond to the drinking question, and data are not missing at random. In other cases, the researcher may have dropped from the analysis some individuals, and the sample selection criteria may be correlated with the outcome of interest or the treatment status.

The model with (potentially) endogenous sample selection considers the case where the dependent count variable y for a given individual is missing if the selection dummy S takes on value zero and is observed if the selection dummy takes on value one. The endogenous treatment and the selection dummies are generated according to a continuous latent variable model:

$$T^* = \mathbf{z}'\gamma + v \quad (\text{A1})$$

$$S^* = \mathbf{r}'\theta + \varphi T + q \quad (\text{A2})$$

with $T = 1(T^* > 0)$, $S = 1(S^* > 0)$, and vectors \mathbf{z} and \mathbf{r} represent a set of explanatory variables (including the constant term) with dimension $K_T \times 1$ and $K_S \times 1$, respectively. γ and θ are conformable vectors of coefficients, φ is the coefficient of the treatment dummy in the selection equation, and v and q are residual terms. We assume that the count y is generated according to the following conditional cumulative distribution function

$$F(y|\eta) \equiv \mathbb{P}(y|\eta) = \begin{cases} \text{not defined} & \text{if } S = 0 \\ \mu^y \exp(-\mu)/y! & \text{if } S = 1 \end{cases} \quad (\text{A3})$$

with

$$y = \begin{cases} \text{missing} & \text{if } S = 0 \\ 0, 1, 2, \dots & \text{if } S = 1 \end{cases} \quad (\text{A4})$$

$$\ln(\mu) = \mathbf{x}'\beta + \delta T + \eta \quad (\text{A5})$$

and all other remaining aspects of the model are the same as in Section 2.1. The main difference between the model presented here and the one in Section 2.1 is the fact that here we use a Poisson distribution for y given $S=1$, whereas we used a zero-truncated Poisson for y given $P=1$ in the endogenous participation model. This is a minor modification that reflects the fact that in the endogenous participation model we considered the $y=0$ count as being generated by a different data generating mechanism from $y>0$ counts.

Another important difference is the fact that now only individuals with $S=1$ in the sample will contribute a non-missing observation for y . The log-likelihood function is now written as follows:

$$\begin{aligned} \log(L) = & \sum_{i, S_i=0} \sum_{\tau} \omega_{\tau} \ln \left\{ \int \mathbb{P}_S(0|\eta) \mathbb{P}_T(\tau|\eta) \phi(\eta) d\eta \right\} \\ & + \sum_{i, S_i=1} \sum_{\tau} \omega_{\tau} \ln \left\{ \int \mathbb{P}_S(1|\eta) \mathbb{P}_T(\tau|\eta) F(y|\eta) \phi(\eta) d\eta \right\} \end{aligned} \quad (\text{A6})$$

We label this model as the SSET-Poisson model. Again, the model can be estimated using MSL.

APPENDIX B. MONTE CARLO SIMULATION STUDY

In this appendix, we investigate the finite sample performance of the EPET-Poisson, the ET-Poisson and the EP-Poisson models in a Monte Carlo simulation study. The aim of the analysis is twofold: (1) to show that the EPET-Poisson estimator can recover the true population parameters when the DGP is indeed EPET-Poisson, and to offer evidence that it is approximately normally distributed in finite samples of moderate size; (2) to show how large the bias is when the researcher fits ET-Poisson or EP-Poisson to data that are truly generated by an EPET-Poisson process and how misleading inference can be if the analyst assumes that the misspecified ET-Poisson/EP-Poisson estimator is asymptotically normally distributed. Given the focus of the study, all experiments discussed here make assumptions that are favorable to the EPET-Poisson estimator.

Simulations are performed using a total sample size of 1500 observations. There are three exogenous variables: x_1 , x_2 and x_3 . These variables are distributed as independent standard normal variates. To facilitate identification, we specify strong exclusion restrictions. Namely: (1) x_1 only enters the equation of T ; (2) x_2 only enters the equation of P ; and (3) x_3 only enters the equation of y . We allow the endogenous treatment T to enter both the participation equation P and the main count response y . The DGP of the simulated data is EPET-Poisson in each replication, with the true population parameters as given in Table BI. In all experiments $\eta \sim N(0, 0.1)$.

The values of the parameters in Table BI and the variance for η were chosen so that the resulting simulated count response variable y will have approximately a mean and a standard deviation (SD) which are near those exhibited by the KT drinking data. Also, these values ensure that approximately 70% of the sample in each replication will have $P = 1$ and $y > 0$ (participants) and 30% will have $P = 0$ and $y = 0$ (non-participants). Again, this was chosen to more or less match the characteristics of the KT drinking data where 21% of the sample have $y = 0$. We decided to have $E(P)$ slightly higher to 0.21 to compensate for the smaller sample size that is used in the simulations. The true population parameters imply that approximately half of the sample is subject to treatment ($T = 1$) in each replication. In the KT drinking data the mean of advice is slightly lower, namely 0.27. Here, again, we set the mean of T to 0.5 to compensate for the smaller sample size used in the simulations. In all experiments the noise/signal ratios are approximately: (a) 0.18 in the T equation; (b) 0.19 in the P equation and (c) 0.16 in the y equation.

We test that the EPET-Poisson estimator can recover the true population parameters regardless the sign that $\rho_{\eta,v}$, $\rho_{\eta,q}$ and $\rho_{v,q}$ take. We set up two groups of experiments (1–4 and 5–8), which only differ in the size of the correlation coefficients. Within each group the only thing that varies is the sign of the aforementioned correlation coefficients:

- Experiment 1: $\rho_{\eta,v} = 0.35$, $\rho_{\eta,q} = 0.35$ and $\rho_{v,q} = 0.13$
- Experiment 2: $\rho_{\eta,v} = -0.35$, $\rho_{\eta,q} = 0.35$ and $\rho_{v,q} = -0.13$
- Experiment 3: $\rho_{\eta,v} = 0.35$, $\rho_{\eta,q} = -0.35$ and $\rho_{v,q} = -0.13$
- Experiment 4: $\rho_{\eta,v} = -0.35$, $\rho_{\eta,q} = -0.35$ and $\rho_{v,q} = 0.13$
- Experiment 5: $\rho_{\eta,v} = 0.78$, $\rho_{\eta,q} = 0.78$ and $\rho_{v,q} = 0.61$
- Experiment 6: $\rho_{\eta,v} = -0.78$, $\rho_{\eta,q} = 0.78$ and $\rho_{v,q} = -0.61$
- Experiment 7: $\rho_{\eta,v} = 0.78$, $\rho_{\eta,q} = -0.78$ and $\rho_{v,q} = -0.61$
- Experiment 8: $\rho_{\eta,v} = -0.78$, $\rho_{\eta,q} = -0.78$ and $\rho_{v,q} = 0.61$

Table BI. Monte Carlo simulations: true population parameters

Variable/equation	T	P	y
Constant	0	0.5	1.56
T	—	1.5	1
x_1	2.5	0	0
x_2	0	-1.9	0
x_3	0	0	-0.8

MSL estimators are asymptotically equivalent to Maximum Likelihood estimators as long as R , the number of draws used to evaluate the simulated likelihood, grows at a faster rate than the square-root of the sample size \sqrt{N} (Gourieroux and Monfort, 1993). Simulation delivers an unbiased approximation of the likelihood $L(\cdot)$. However, MSL maximizes $\ln[\hat{L}(\cdot)]$ rather than $\hat{L}(\cdot)$, and simulation of $\ln[\hat{L}(\cdot)]$ does not give an unbiased estimator for $\ln[L(\cdot)]$. This bias does not affect the consistency of the MSL estimator. However, to achieve asymptotic normality it is required that $\sqrt{N}/R \rightarrow 0$. If this condition does not hold the covariance matrix estimator of MSL is incorrect; see Cameron and Trivedi (2005, p. 394–396) and Train (2003, p. 258–259). For this reason the analyst is advised to use a *large enough* number of draws R . There is no general rule to choose R and how large it should be depends on particular applications. In the case of the EPET-Poisson, preliminary Monte Carlo simulation experiments with 200 replications showed that MSL estimators using less than 1000 Halton draws deliver slightly underestimated standard errors and that the coverage of the parameters was below the advertised 95%. Once the number of Halton draws R was set to values larger than 1000 the EPET-Poisson behaved well.

Table BII and BIII present Monte Carlo simulation results for experiments 1 to 4. In all these experiments we used 1600 Halton draws. Table BII reports the performance of EPET-Poisson, the ET-Poisson and the EP-Poisson in terms of bias and Monte Carlo SD. Results show that, as expected, the EPET-Poisson delivers largely unbiased estimators of all the parameters. Indeed, for all parameters, percentage bias is below 3.1%. In all four experiments the computed bias for the coefficient of main interest b_{yT} (i.e. the coefficient of the treatment T on the main count response y) is not statistically different from zero at a significance level of 5%.

In the case of the ET-Poisson model the reported bias is large with respect to the size of the parameters. To give a relative idea of the size of the bias, the reader can consider that the percentage bias of b_{yT} in ET-Poisson is 76% in experiment 1 and 78% in experiment 4. This bias is significantly different from zero at 5% in all the four experiments. The EP-Poisson is doing better in this respect, and the percentage bias of b_{yT} is 7% in experiment 1 and -7% in experiment 4.

Table BIII shows the performance of the different models in terms of standard error and the coverage achieved. From this table the reader can see that the EPET-Poisson estimator delivers an average standard error that is broadly equal to the Monte Carlo SD. In other words, the finite sample variation of the EPET-Poisson estimator is broadly the same as the variation one expects under the

Table BII. Monte Carlo simulations: bias and SD

Parameters	$\rho_{\eta,v} = 0.35$ $\rho_{\eta,q} = 0.35$ $\rho_{v,q} = 0.13$		$\rho_{\eta,v} = -0.35$ $\rho_{\eta,q} = 0.35$ $\rho_{v,q} = -0.13$		$\rho_{\eta,v} = 0.35$ $\rho_{\eta,q} = -0.35$ $\rho_{v,q} = -0.13$		$\rho_{\eta,v} = -0.35$ $\rho_{\eta,q} = -0.35$ $\rho_{v,q} = 0.13$	
	Bias	SD	Bias	SD	Bias	SD	Bias	SD
EPET-Poisson								
$\hat{\rho}_{\eta,v}$	0.0036	0.0967	0.0047	0.0978	0.0043	0.0979	0.0057	0.0998
$\hat{\rho}_{\eta,q}$	0.0040	0.1333	-0.0039	0.1311	-0.0013	0.1247	-0.0045	0.1292
$\hat{\rho}_{v,q}$	0.0012	0.0567	-0.0004	0.0563	0.0005	0.0555	-0.0009	0.0553
$\hat{\rho}_{PT}$	0.0136	0.1209	0.0018	0.1210	0.0071	0.1174	0.0068	0.1211
\hat{b}_{yT}	0.0008	0.0390	0.0016	0.0383	0.0004	0.0392	0.0031	0.0410
ET-Poisson								
$\hat{\rho}_{\eta,q}$	-0.3257	0.0636	-0.1345	0.0618	-0.4200	0.0633	-0.2185	0.0641
\hat{b}_{yT}	0.7600	0.0657	0.7603	0.0681	0.7777	0.0643	0.7786	0.0647
EP-Poisson								
$\hat{\rho}_{\eta,q}$	0.0188	0.1392	-0.0200	0.1378	-0.0176	0.1340	0.0089	0.1349
\hat{b}_{PT}	0.0891	0.1157	-0.0703	0.1165	-0.0656	0.1131	0.0810	0.1176
\hat{b}_{yT}	0.0712	0.0344	-0.0683	0.0332	0.0707	0.0349	-0.0675	0.0354

Note: Monte Carlo simulation study with 1000 replications, $N = 1500$ and 1600 Halton draws. SD ≡ Monte Carlo standard deviation.

Table BIII. Monte Carlo simulations: ASE/SD and nominal coverage

	$\hat{\rho}_{\eta,v} = 0.35$ $\hat{\rho}_{\eta,q} = 0.35$ $\hat{\rho}_{v,q} = 0.13$		$\hat{\rho}_{\eta,v} = -0.35$ $\hat{\rho}_{\eta,q} = 0.35$ $\hat{\rho}_{v,q} = -0.13$		$\hat{\rho}_{\eta,v} = 0.35$ $\hat{\rho}_{\eta,q} = -0.35$ $\hat{\rho}_{v,q} = -0.13$		$\hat{\rho}_{\eta,v} = -0.35$ $\hat{\rho}_{\eta,q} = -0.35$ $\hat{\rho}_{v,q} = 0.13$	
Parameters	ASE/SD	Covr	ASE/SD	Covr	ASE/SD	Covr	ASE/SD	Covr
EPET-Poisson								
$\hat{\rho}_{\eta,v}$	1.041	95.4	1.015	95.1	1.033	95.8	1.010	95.0
$\hat{\rho}_{\eta,q}$	0.992	93.6	0.996	94.4	1.019	94.2	0.990	93.9
$\hat{\rho}_{v,q}$	1.022	92.9	1.032	92.5	1.039	94.2	1.023	94.0
\hat{b}_{PT}	1.014	94.9	0.971	94.5	1.003	95.6	1.010	95.4
\hat{b}_{yT}	1.012	94.8	0.994	94.3	1.001	95.0	0.968	94.0
ET-Poisson								
$\hat{\rho}_{\eta,q}$	1.109	0.02	1.139	53.5	1.123	0	1.131	11.6
\hat{b}_{yT}	1.623	0	1.607	0	1.648	0	1.634	0
EP-Poisson								
$\hat{\rho}_{\eta,q}$	1.014	93.5	1.024	94.4	1.024	94.4	1.001	93.6
\hat{b}_{PT}	1.011	89.3	0.971	89.3	1.000	89.7	0.996	89.9
\hat{b}_{yT}	0.983	44.6	0.988	45.0	0.976	44.7	0.959	48.7

Note. ASE/SD=(average standard error/Monte Carlo standard deviation). Covr=Coverage (%).

assumption that the EPET-Poisson estimator is normally distributed. This last observation is also supported by the fact that coverage of all parameters estimated by the EPET-Poisson achieves approximately the advertised size of 95%. This is true across all four experiments. In conclusion, there is strong evidence that EPET-Poisson is asymptotically normally distributed as expected.

Performance of the ET-Poisson in terms of standard error and coverage is disappointing. Standard errors are largely underestimated and coverage is far from the advertised 95%. In fact, the true value of the population parameter for the main coefficient of interest, b_{yT} , lied outside the 95% confidence interval in every single replication and across all four experiments. When the true DGP is EPET-Poisson also the EP-Poisson model performs badly in terms of coverage of the main parameters of interest.

Experiments 5–8 (see Tables BIV and BV) give similar results and for the sake of brevity we do not add any further comment on those experiments.²⁸ Summarizing, we find that if the DGP is EPET-Poisson: (1) the EPET-Poisson estimator is consistent and asymptotically normally distributed and (2) fitting a ET-Poisson or a EP-Poisson delivers a severely biased estimator and underestimated standard errors. Hence, performing inference on the basis of ET-Poisson or EP-Poisson when the DGP is EPET-Poisson may lead to severely misleading inference. In all simulations that we reported here, exclusion restrictions hold both in the DGP and in the estimated model. In other words, they refer to the best-case scenario in which the researcher can use (valid) exclusion restrictions. We also run Monte Carlo simulations in which all equations include the same covariates, i.e. exclusion restrictions do not hold neither in the DGP nor in the estimated model, and obtained a very similar performance of the EPET-Poisson estimator, confirming that the model is formally identified even without exclusion restrictions. When the DGP is EPET-Poisson, without exclusion restrictions the bias for the ET-Poisson

²⁸The only thing to note is that the standard errors for $\hat{\rho}_{v,q}$ are ‘slightly’ larger than the calculated Monte Carlo standard deviations. This overestimation, however, is relatively small with respect to the size of $\hat{\rho}_{v,q}$ and it does not affect the coverage in any case, which remains around 94%. Correlation coefficients are special in the sense that having to lie between -1 and 1 , they cannot truly be distributed as normal variates. Also, $\hat{\rho}_{v,q}$ is a function of λ_1 , λ_2 and σ_η and its standard error is calculated using the Delta method. If $\hat{\rho}_{v,q}$, as a function of λ_1 , λ_2 and σ_η is approximately quadratic around the ML estimates, then a second order approximation of the standard errors for $\hat{\rho}_{v,q}$ will be precise. However, if $\hat{\rho}_{v,q}$ is not quadratic enough, then calculation of the standard error of $\hat{\rho}_{v,q}$ by the Delta method may be slightly off target.

Table BIV. Monte Carlo simulations: bias and SD

	$\hat{\rho}_{\eta,v} = 0.78$ $\hat{\rho}_{\eta,q} = 0.78$ $\hat{\rho}_{v,q} = 0.61$		$\hat{\rho}_{\eta,v} = -0.78$ $\hat{\rho}_{\eta,q} = 0.78$ $\hat{\rho}_{v,q} = -0.61$		$\hat{\rho}_{\eta,v} = 0.78$ $\hat{\rho}_{\eta,q} = -0.78$ $\hat{\rho}_{v,q} = -0.61$		$\hat{\rho}_{\eta,v} = -0.78$ $\hat{\rho}_{\eta,q} = -0.78$ $\hat{\rho}_{v,q} = 0.61$	
Parameters	Bias	SD	Bias	SD	Bias	SD	Bias	SD
EPET-Poisson								
$\hat{\rho}_{\eta,v}$	0.0027	0.0667	0.0046	0.0635	0.0026	0.0683	0.0053	0.0666
$\hat{\rho}_{\eta,q}$	-0.0024	0.0790	-0.0011	0.0797	0.0004	0.0743	-0.0016	0.0764
$\hat{\rho}_{v,q}$	-0.0012	0.0699	0.0016	0.0709	0.0008	0.0663	0.0014	0.0669
\hat{b}_{PT}	0.0053	0.1345	0.0006	0.1122	0.0108	0.1104	0.0073	0.1352
\hat{b}_{yT}	-0.0014	0.0364	0.0001	0.0344	-0.0005	0.0364	0.0005	0.0384
ET-Poisson								
$\hat{\rho}_{\eta,q}$	-0.4849	0.0602	-0.3051	0.0589	-0.9426	0.0647	-0.7292	0.0668
\hat{b}_{yT}	0.7398	0.0663	0.7504	0.0679	0.7904	0.0593	0.7908	0.0616
EP-Poisson								
$\hat{\rho}_{\eta,q}$	0.0087	0.0995	-0.0286	0.1030	0.0257	0.0948	-0.0127	0.0916
\hat{b}_{PT}	0.3979	0.1234	-0.3402	0.1087	-0.3292	0.1068	0.4032	0.1255
\hat{b}_{yT}	0.1527	0.0320	-0.1523	0.0309	0.1520	0.0322	-0.1545	0.0353

Note: Monte Carlo simulation study with 1000 replications, $N = 1500$, and 1600 Halton draws. SD≡Monte Carlo standard deviation.

Table BV. Monte Carlo simulations: ASE/SD and nominal coverage

	$\hat{\rho}_{\eta,v} = 0.78$ $\hat{\rho}_{\eta,q} = 0.78$ $\hat{\rho}_{v,q} = 0.61$		$\hat{\rho}_{\eta,v} = -0.78$ $\hat{\rho}_{\eta,q} = 0.78$ $\hat{\rho}_{v,q} = -0.61$		$\hat{\rho}_{\eta,v} = 0.78$ $\hat{\rho}_{\eta,q} = -0.78$ $\hat{\rho}_{v,q} = -0.61$		$\hat{\rho}_{\eta,v} = -0.78$ $\hat{\rho}_{\eta,q} = -0.78$ $\hat{\rho}_{v,q} = 0.61$	
Parameters	ASE/SD	Covr	ASE/SD	Covr	ASE/SD	Covr	ASE/SD	Covr
EPET-Poisson								
$\hat{\rho}_{\eta,v}$	0.983	93.8	1.004	94.1	0.970	93.0	0.989	93.3
$\hat{\rho}_{\eta,q}$	1.036	94.9	0.999	94.4	1.025	94.7	1.021	95.8
$\hat{\rho}_{v,q}$	2.337	94.1	7.157	92.8	3.472	92.9	2.523	93.0
\hat{b}_{PT}	1.006	95.3	0.981	95.3	1.007	94.3	0.994	95.6
\hat{b}_{yT}	0.997	95.5	0.988	94.8	0.984	95.3	0.965	93.7
ET-Poisson								
$\hat{\rho}_{\eta,q}$	1.076	0	1.047	0	1.079	0	1.123	0
\hat{b}_{yT}	1.546	0	1.650	0	1.818	0	1.670	0
EP-Poisson								
$\hat{\rho}_{\eta,q}$	1.065	95.4	1.040	95.7	1.057	96.1	1.062	93.6
\hat{b}_{PT}	1.009	7.5	0.979	12.1	1.000	14.7	0.993	8.1
\hat{b}_{yT}	1.036	0	0.986	0.2	1.013	0.2	0.952	7.8

Note: ASE/SD≡(average standard error/Monte Carlo standard deviation). Covr≡Coverage (%).

or the EP-Poisson models tends to increase, especially when the correlation between the error terms is high. These additional results are available from the corresponding author upon request.

ACKNOWLEDGEMENTS

Both authors equally contributed to this paper. Participants to presentations held at the 19th European Workshop on Econometrics and Health Economics (Lausanne), the National Autonomous

University of México and the College of México are acknowledged for their suggestions. We are grateful to two anonymous referees, João Santos Silva and Stefanie Schurer for their useful comments. The usual disclaimers apply.

REFERENCES

- Cameron A, Trivedi P. 2005. *Microeconometrics: Methods and Applications*. Cambridge University Press: New York.
- Carneiro P, Hansen K, Heckman J. 2003. Estimating distributions of treatment effects with an application to the returns to schooling and measurement of the effects of uncertainty on college choice. *International Economic Review* **44**(2): 361–422.
- Chamberlain G. 1980. Analysis of covariance with qualitative data. *The Review of Economic Studies* **47**(1): 225–238.
- Cox D, Wermuth N. 1996. *Multivariate Dependencies: Models, Analysis and Interpretation*. Monographs on Statistics and Applied Probability, vol. 67. Chapman & Hall: London.
- Cragg J. 1971. Some statistical models for limited dependent variables with application to the demand for durable goods. *Econometrica* **39**: 829–844.
- Deb P, Trivedi P. 2006. Specification and simulated likelihood estimation of a non-normal treatment-outcome model with selection: applications to health care utilization. *Econometrics Journal* **9**(2): 307–331.
- Gourieroux C. 2000. *Econometrics of Qualitative Dependent Variables*. Cambridge University Press: New York.
- Gourieroux C, Monfort A. 1993. Simulation-based inference: a survey with special reference to panel data models. *Journal of Econometrics* **59**(1–2): 5–33.
- Greene W. 1997. *FIML Estimation of Sample Selection Models for Count Data*. Stern School of Business: New York University: New York. Manuscript.
- Greene W. 2009. Models for count data with endogenous participation. *Empirical Economics* **36**(1): 133–173.
- Jones A. 1989. A double-hurdle model for cigarette consumption. *Journal of Applied Econometrics* **4**(1): 23–39.
- Keane M. 1992. A note on identification in the multinomial probit model. *Journal of Business and Economic Statistics* **10**(2): 192–200.
- Kenkel D, Terza J. 2001. The effect of physician advice on alcohol consumption: count regression with an endogenous treatment effect. *Journal of Applied Econometrics* **16**(2): 165–184.
- Li Q, Trivedi P. 2009. Impact of prescription drug coverage on drug expenditure of the elderly – evidence from a two-part model with endogeneity. Mimeo. (Available from: http://mypage.iu.edu/~trivedi/drug_all_0609.pdf.)
- Melkersson M, Rooth D. 2000. Modeling female fertility using inflated count data models. *Journal of Population Economics* **13**(2): 189–203.
- Miranda A. 2004. FIML estimation of an endogenous switching model for count data. *The Stata Journal* **4**(1): 40–49.
- Miranda A, Rabe-Hesketh S. 2006. Maximum likelihood estimation of endogenous switching and sample selection models for binary, ordinal, and count variables. *The Stata Journal* **6**(3): 285–308.
- Monfardini C, Radice R. 2008. Testing exogeneity in the bivariate probit model: a Monte Carlo study. *Oxford Bulletin of Economics and Statistics* **70**(2): 271–282.
- Mullahy J. 1986. Specification and testing of some modified count data models. *Journal of Econometrics* **33**(3): 341–365.
- Mullahy J. 1998. Much ado about two: reconsidering retransformation and the two-part model in health econometrics. *Journal of Health Economics* **17**(3): 247–281.
- Mundlak Y. 1978. On the pooling of time series and cross section data. *Econometrica* **46**(1): 69–85.
- Puhani P. 2000. The Heckman correction for sample selection and its critique. *Journal of Economic Surveys* **14**(1): 53–68.
- Riphahn R, Wambach A, Million A. 2003. Incentive effects in the demand for health care: a bivariate panel count data estimation. *Journal of Applied Econometrics* **4**(18): 387–405.
- Terza JV. 1985. A tobit-type estimator for the censored poisson regression model. *Economics Letters* **18**(4): 361–365.
- Terza JV. 1998. Estimating count data models with endogenous switching: sample selection and endogenous treatment effects. *Journal of Econometrics* **84**(4): 129–154.
- Terza JV. 2009. Parametric nonlinear regression with endogenous switching. *Econometric Reviews* **28**(6): 555–580.
- Terza JV, Kenkel DS, Lin T-F, Sakata S. 2008. Care-giver advice as a preventive measure for drinking during pregnancy: zeros, categorical outcome responses, and endogeneity. *Health Economics* **17**(1): 41–54.
- Train K. 2003. *Discrete Choice Methods With Simulation*. Cambridge University Press: Cambridge (UK).
- Windmeijer F, Santos Silva J. 1997. Endogeneity in count data models: an application to demand for health care. *Journal of Applied Econometrics* **12**(3): 281–294.
- Winkelmann R. 1998. Count data models with selectivity. *Econometric Reviews* **17**(4): 339–359.
- Winkelmann R. 2008. *Econometric Analysis of Count Data* (5th edn). Springer: Berlin, Heidelberg, New York.